

Special Statistical Challenges and Opportunities in EMA: A Tale of Frustration

Joe Schwartz

September 5, 2003
Charleston, S.C.



Characteristics of Real-time Data

- A moderate to large number of observations for each person
- Unequal numbers of observations across persons (differences in sleep, missing data, etc)
- Serial autocorrelation of observations within-persons
- Differences between persons in average levels

What if you're primarily interested in analyzing people's average levels?

- Could just compute each person's average and use this as an outcome measure

or

- Could use multilevel (mixed) models
 - a general class of models that represents an extension of traditional repeated measures ANOVA to unbalanced data

Advantage of Multilevel Models When Analyzing Person Differences

- Multilevel models treat person means as a “latent variable”
 - Computed means are differentially reliable, varying primarily according to the number of assessments available for each subject
 - Multilevel models give somewhat greater weight to those averages that are based on more assessments
 - If you have many observations for all persons or roughly equal numbers of observations per person and you are NOT interested in within-person relationships, then multilevel modeling probably offers little advantage
 - Allows one to more easily statistically control for time-varying “third” factors.

Q1: Does Average Frustration Vary Significantly Across People?

- If we compute person means
 - N=68, Mean=34.7, StdDev=21.7
 - No test of significance
- Nested ANOVA (estimates of between-person Var)
 - PROC NESTED: StdDev=21.9, no test of signif
 - PROC VARCOMP:
 - Method=Type I: same as PROC NESTED
 - Method=ML: StdDev=21.4, no test of signif

Q1: Does Average Frustration Vary Significantly Across People?

- Multilevel model with no predictors
 - Mean (of “true” person means) = 34.7
 - StdDev (of “true” person means) = 21.4, $p < .0001$
 - StdDev of w/in person deviations = 20.6, $p < .0001$
- Intraclass correlation (ICC)
 - Ratio of Between-person Variance / Total Variance

Answer to Q1

Yes, there are significant differences among persons in mean level of frustration.

Slightly more than half the variance of frustration is due to between-person differences in mean levels; the rest is due to within-person fluctuations in frustration (including measurement error).

Decomposition of Variance for Frustration - traditional methods

N = (5321, 68)

Source of Variance	Person Means	Nested ANOVA	
		Least Sq	Max Like
Between-person	472	481	458
Within-person	---	424	424
Total		905	882
%Between-person*	---	53%	52%

* Intraclass Correlation (ICC)

Decomposition of Variance for Frustration - multilevel models

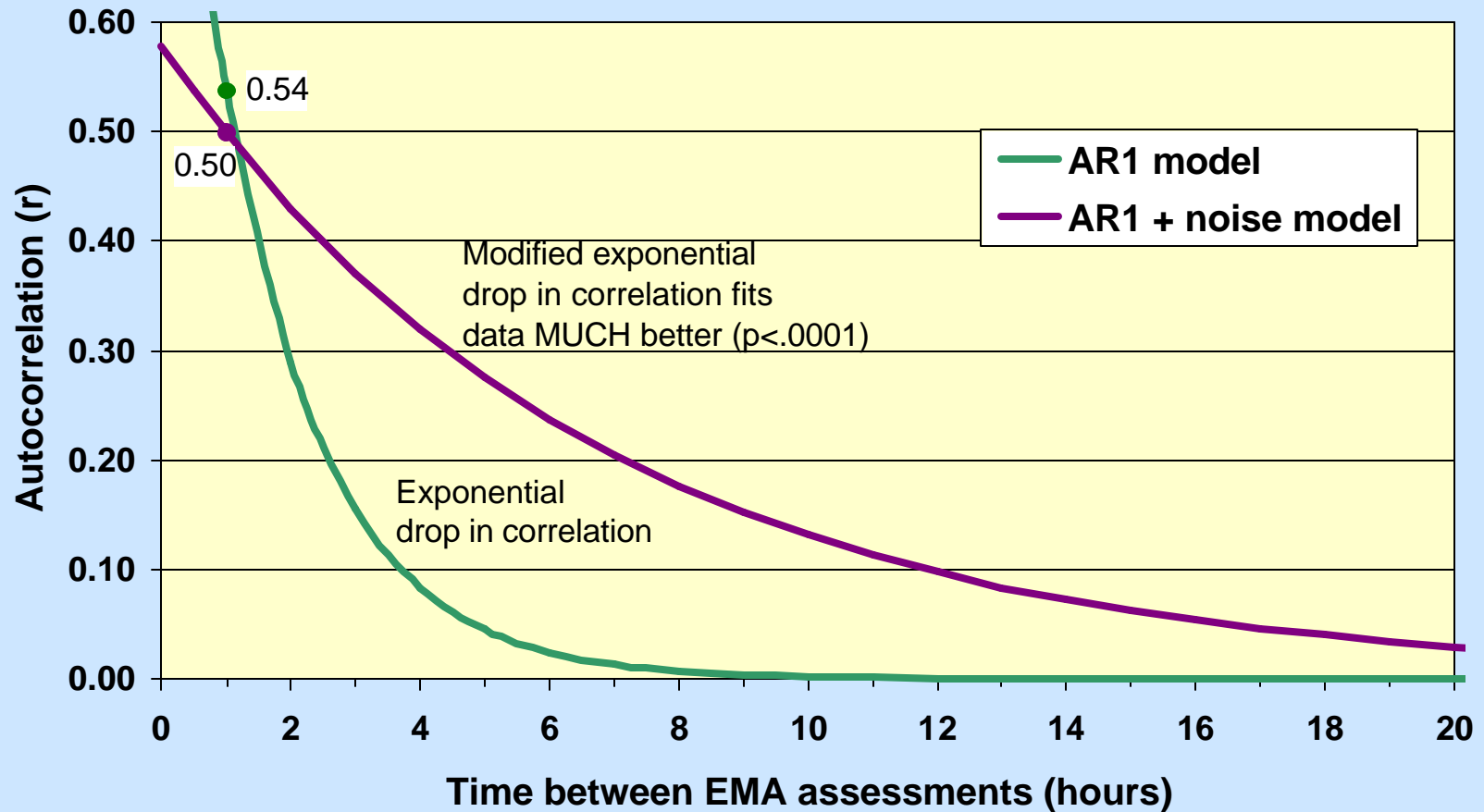
N = (5321, 68)

Source of Variance	Max Like	+autocor	+ "noise"
Between-person	458*	450*	436*
Within-person	424*	441*	445*
State Var			257*
Autocor (1 hr)		0.54*	0.86*
Error Var			187*
%Between-person ^a	52%	50%	50%
-2 LL	47589	46905	46686

^a Intraclass Correlation (ICC)

* p<.0001

Variogram: Estimates from Two Models of Within-Person Autocorrelation of Frustration (AR1 and "AR1 + Random Noise")



Advantages of Multilevel Models over Traditional Repeated Measures ANOVA

- Multilevel models can handle unequal numbers of observations for each subject
- Multilevel models can handle a wider variety of correlated error patterns
 - Standard repeated measures ANOVA procedures allow for compound symmetry, Greenhouse-Geiser, Huynh-Feldt, and MANOVA error models
 - Given the time-series nature of most Real-Time data, residuals will usually exhibit serial autocorrelation

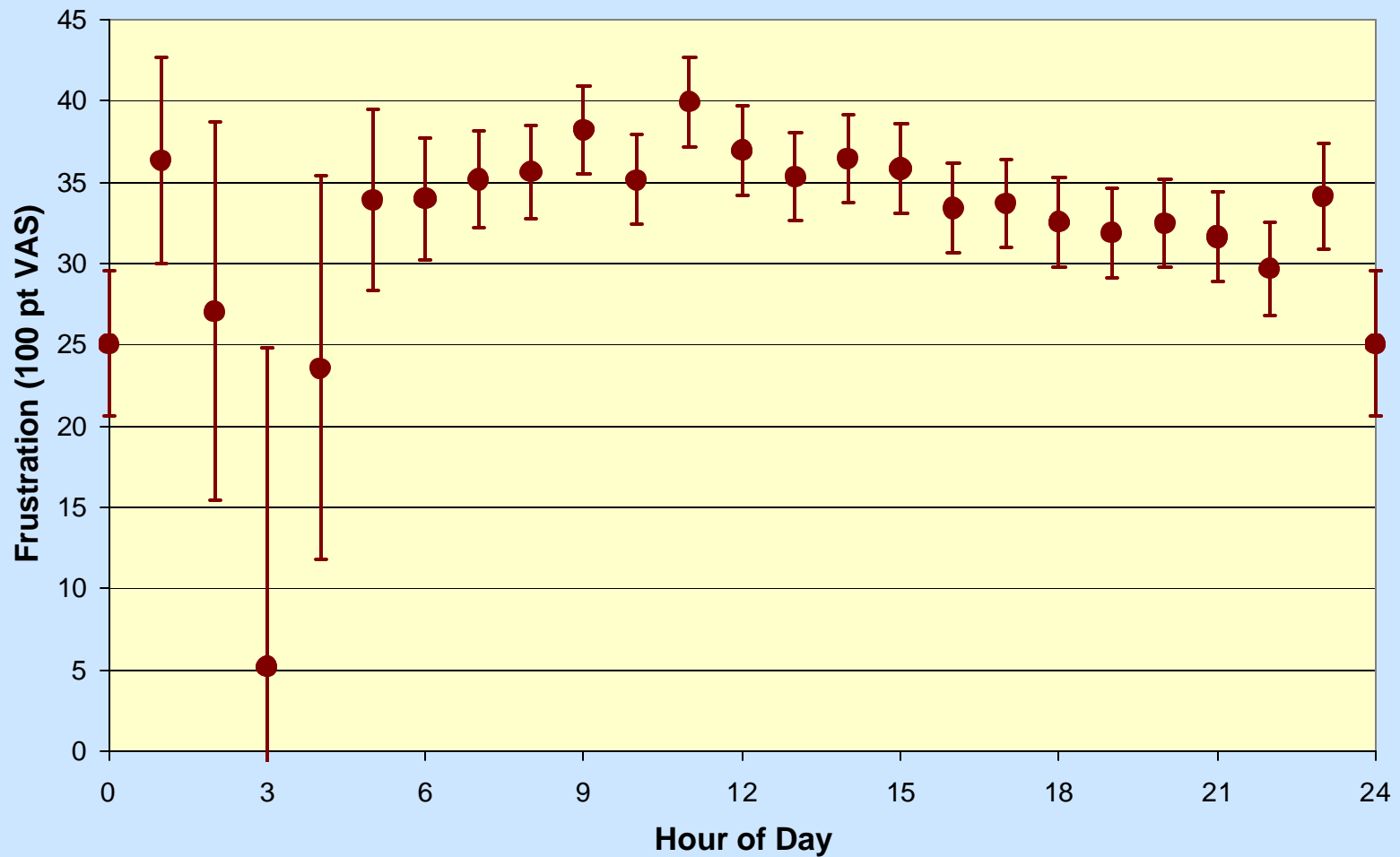
Advantages of Multilevel Models over Traditional Repeated Measures ANOVA

- Traditional repeated measures ANOVA requires balanced within-person designs; multilevel models handle unbalanced within-person factors (similar to regression analysis), allowing one to analyze or statistically control for a wider range of within-person factors (including continuous variables)

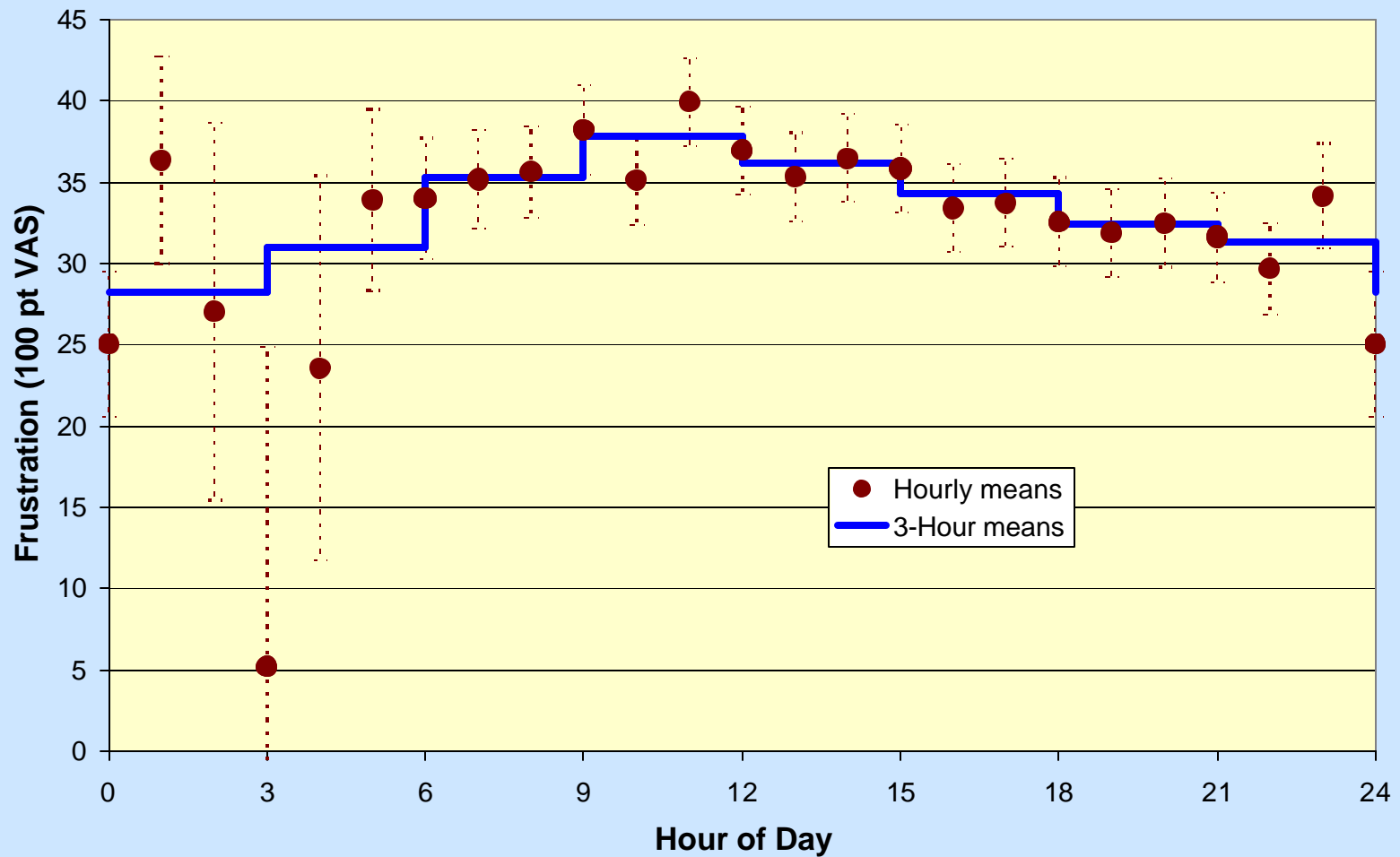
Q2: Are There Temporal Patterns to Frustration?

- Time-of-day effects
 - hour of day, 3-hour blocks
 - diurnal cycle, circadian rhythm
 - sinusoidal curve

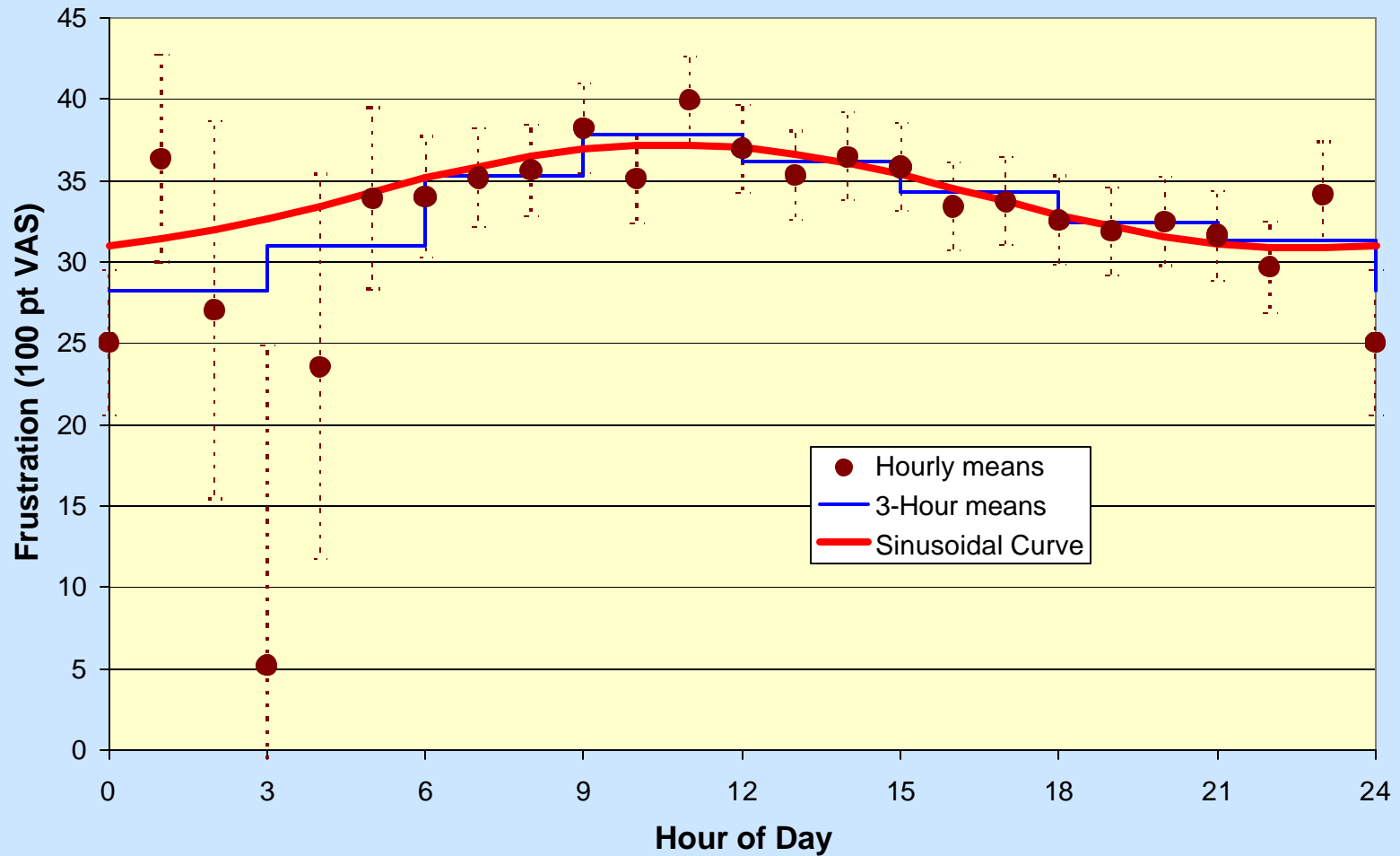
Diurnal Pattern of Self-reported Frustration (hourly means ± 1 se)



Diurnal Pattern of Self-reported Frustration (1-hour and 3-hour means)



Diurnal Pattern of Self-reported Frustration (1-hour and 3-hour means, sinusoidal curve)



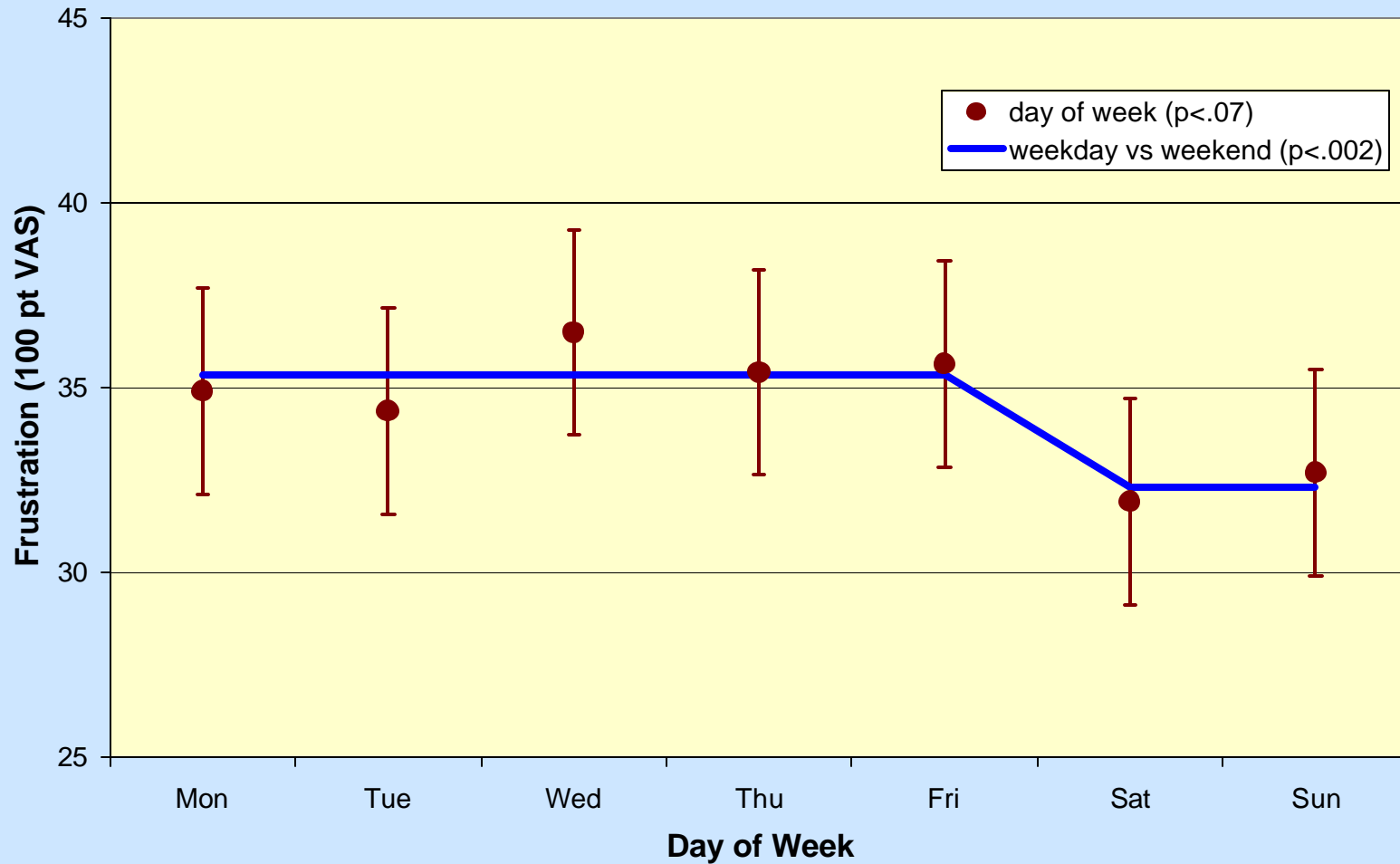
Comparative Fit of 3 Models of Diurnal Pattern

Model	DF	Deviance (-2 LL)	Δ Dev	p
A. Hourly means	23	46602.8	83.0	.0001
B. 3-Hour means	7	46631.5	54.3	.0001
C. Sinusoidal curve	2	46634.1	51.7	.0001
D. No diurnal pattern	0	46685.8		
Model A vs Model B	16		28.7	.026
Model A vs Model C	21		31.3	.069
Model B vs Model C	5		2.6	.76

Q2: Are There Temporal Patterns to Frustration?

- Time-of-day effects
 - hour of day, 3-hour blocks
 - diurnal cycle, circadian rhythm
 - sinusoidal curve
- Day-of-week effects
 - weekday vs weekend

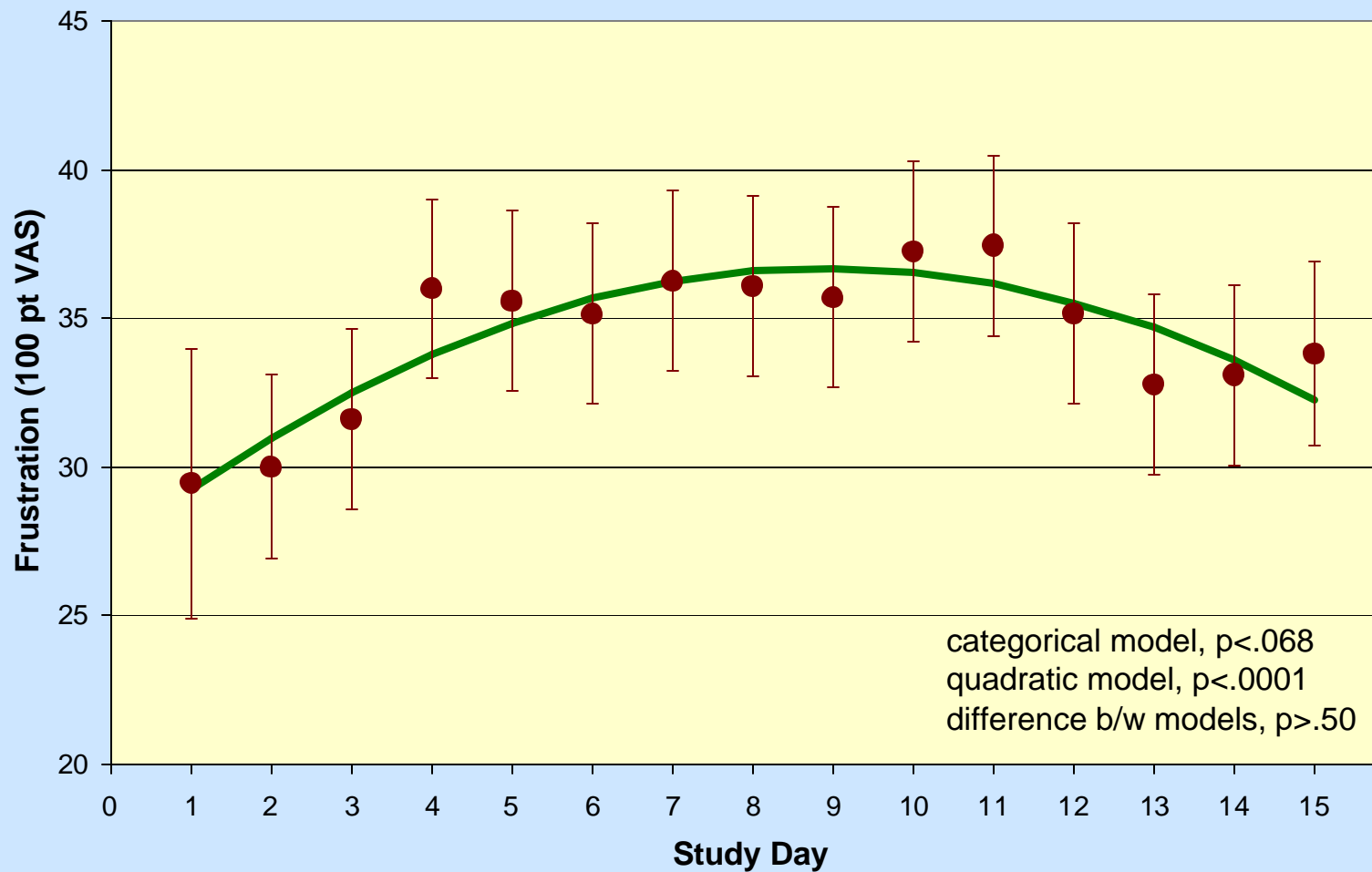
Weekly Pattern of Self-reported Frustration (day of week and weekday vs weekend)



Q2: Are There Temporal Patterns to Frustration?

- Time-of-day effects
 - hour of day, 3-hour blocks
 - diurnal cycle, circadian rhythm
 - sinusoidal curve
- Day-of-week effects
 - weekday vs weekend
- Study day effects

Pattern of Self-reported Frustration Over Study Period



Answer to Q2

Yes, there are temporal patterns.

Frustration tends to peak mid-day and be lower at night and in the early morning. The average diurnal cycle is adequately approximated by a sinusoidal curve.

Frustration is lower during the weekend than during weekdays.

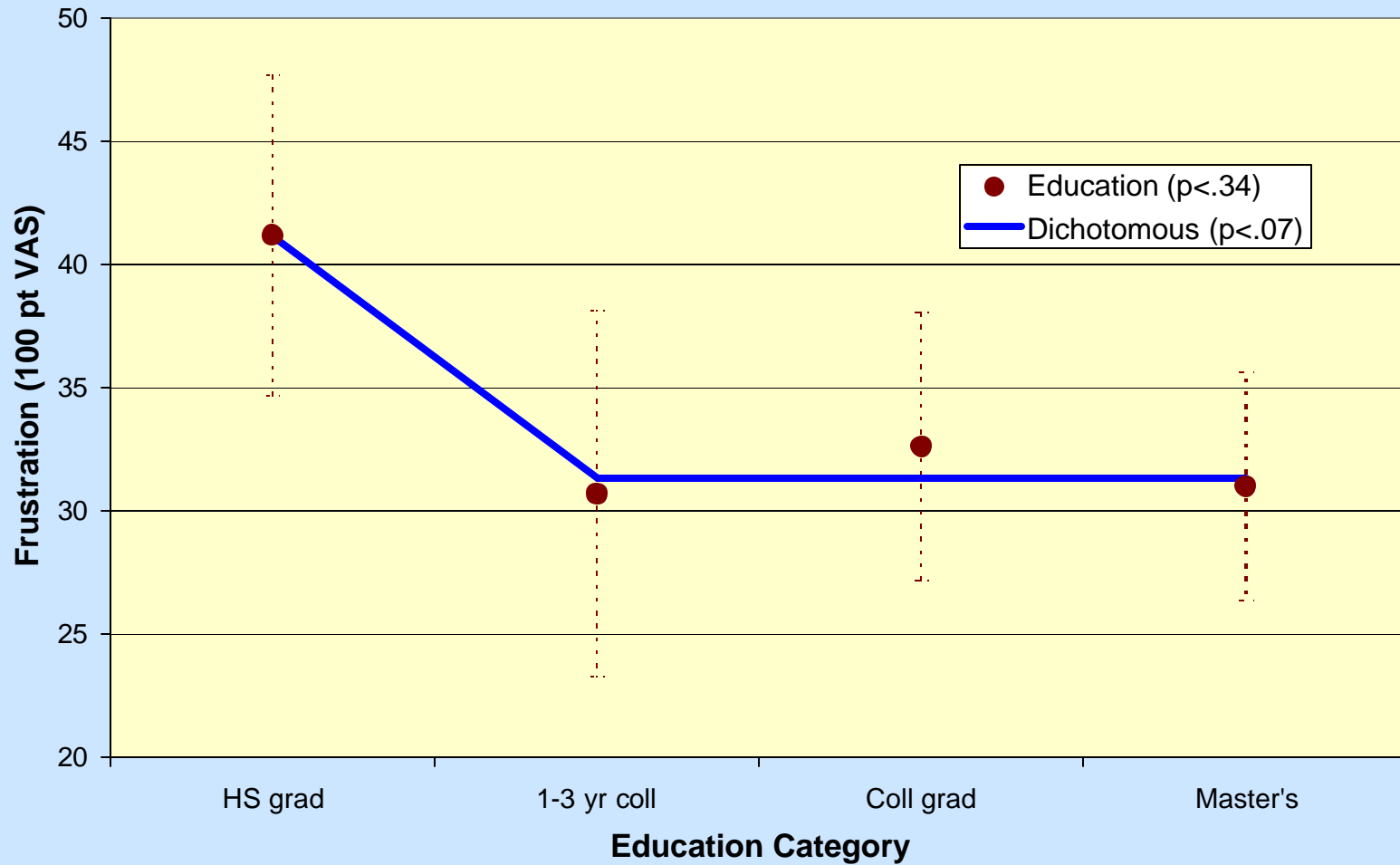
Over the two weeks of data collection, average levels of frustration increased and then decreased (adequately fit with a quadratic curve)

Q3: Does Frustration Vary by Socioeconomic Status (SES)? (between-person predictors)

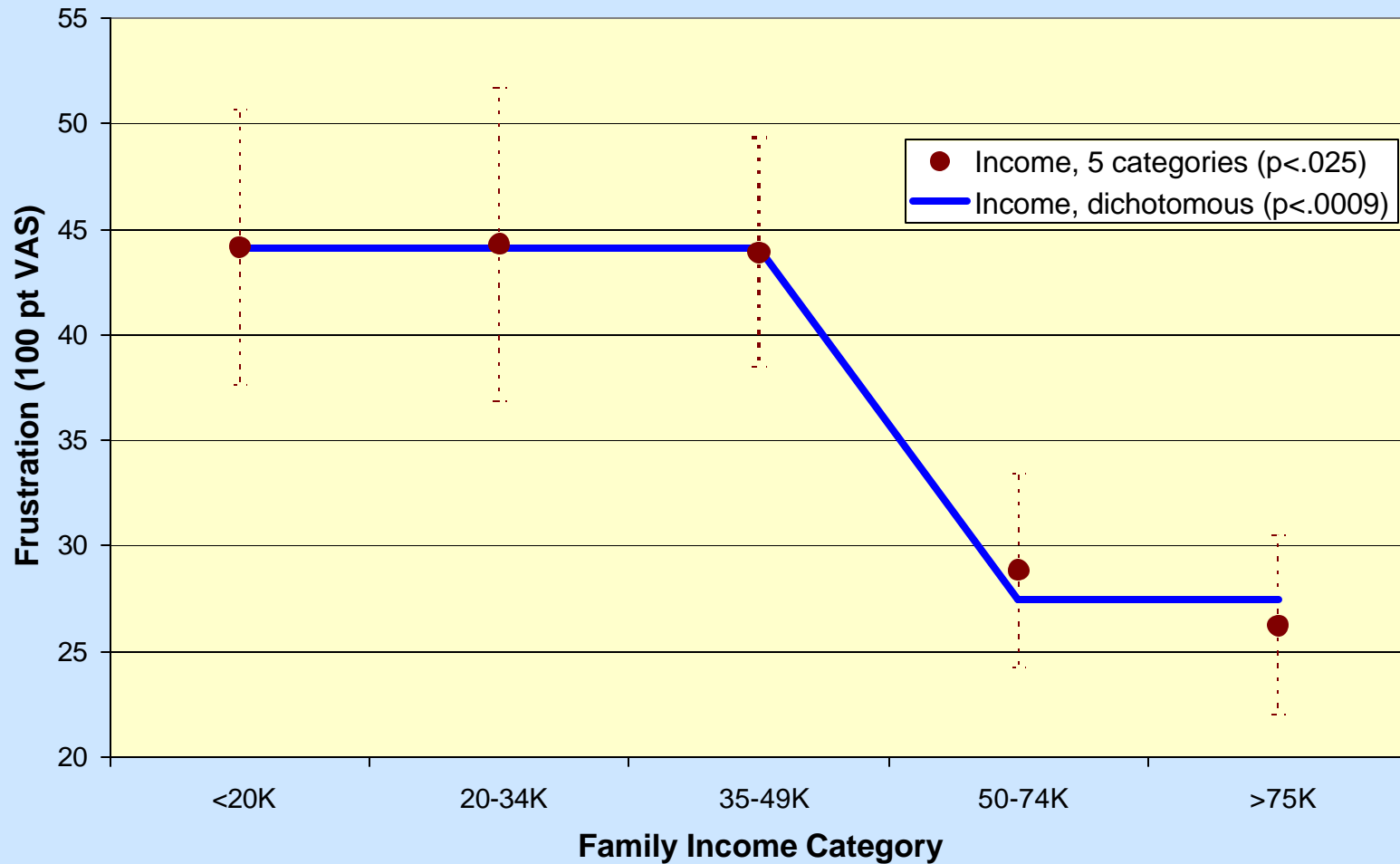
To be more precise,

“Do people’s average levels of frustration differ according to their education and/or income?”

Mean Self-reported Frustration, by Education



Mean Self-reported Frustration, by Family Income



Answer to Q3

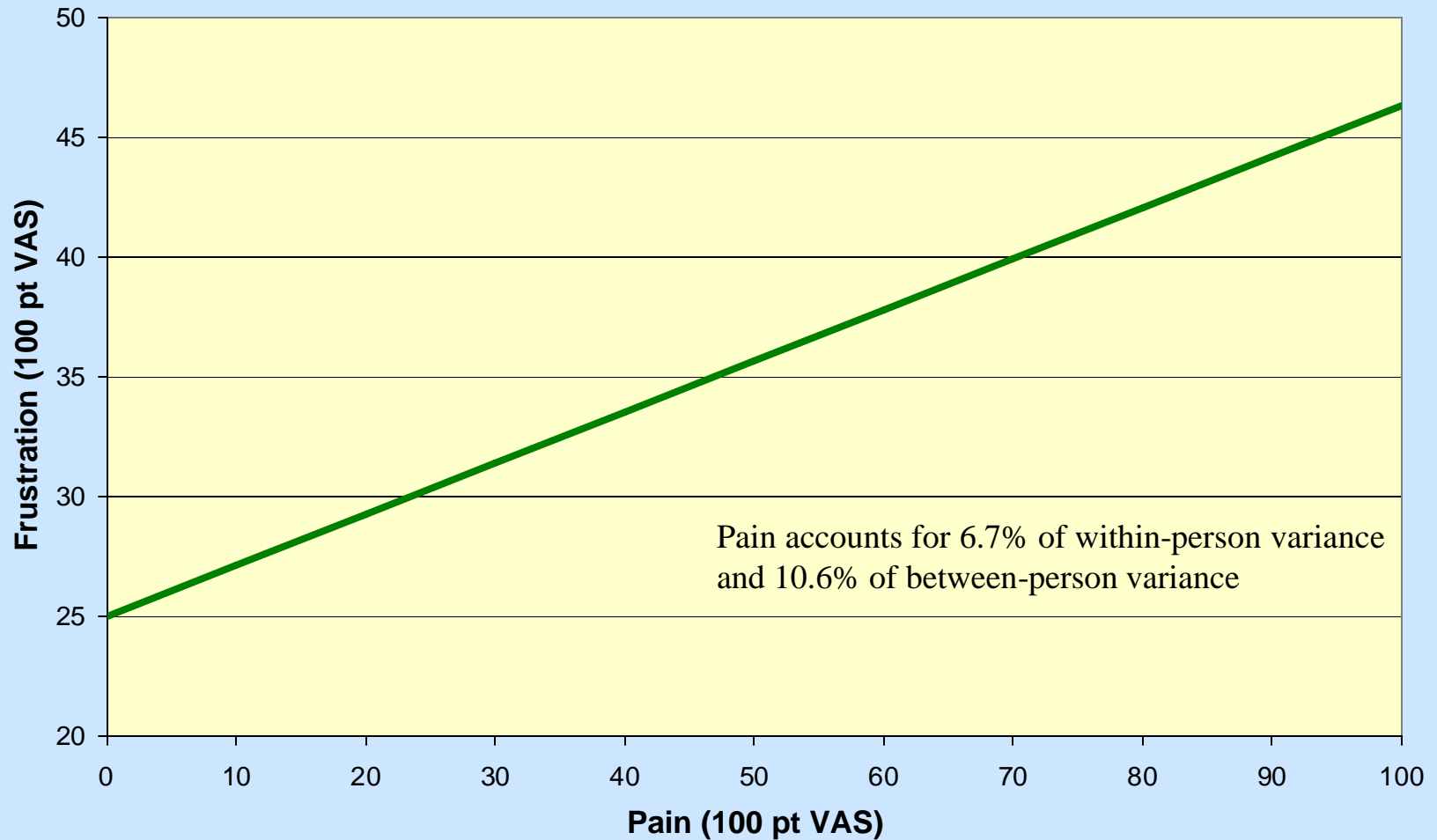
Yes, those with lower family incomes have higher average levels of frustration. Inverse relationship with education is marginally significant.

Q4: Is Frustration Related to Pain? (a within-person predictor)

To be more precise,

“Are changes in frustration associated with
changes in pain?”

Within-Person Relationship of Frustration to Pain (assuming relationship is the same for all people)

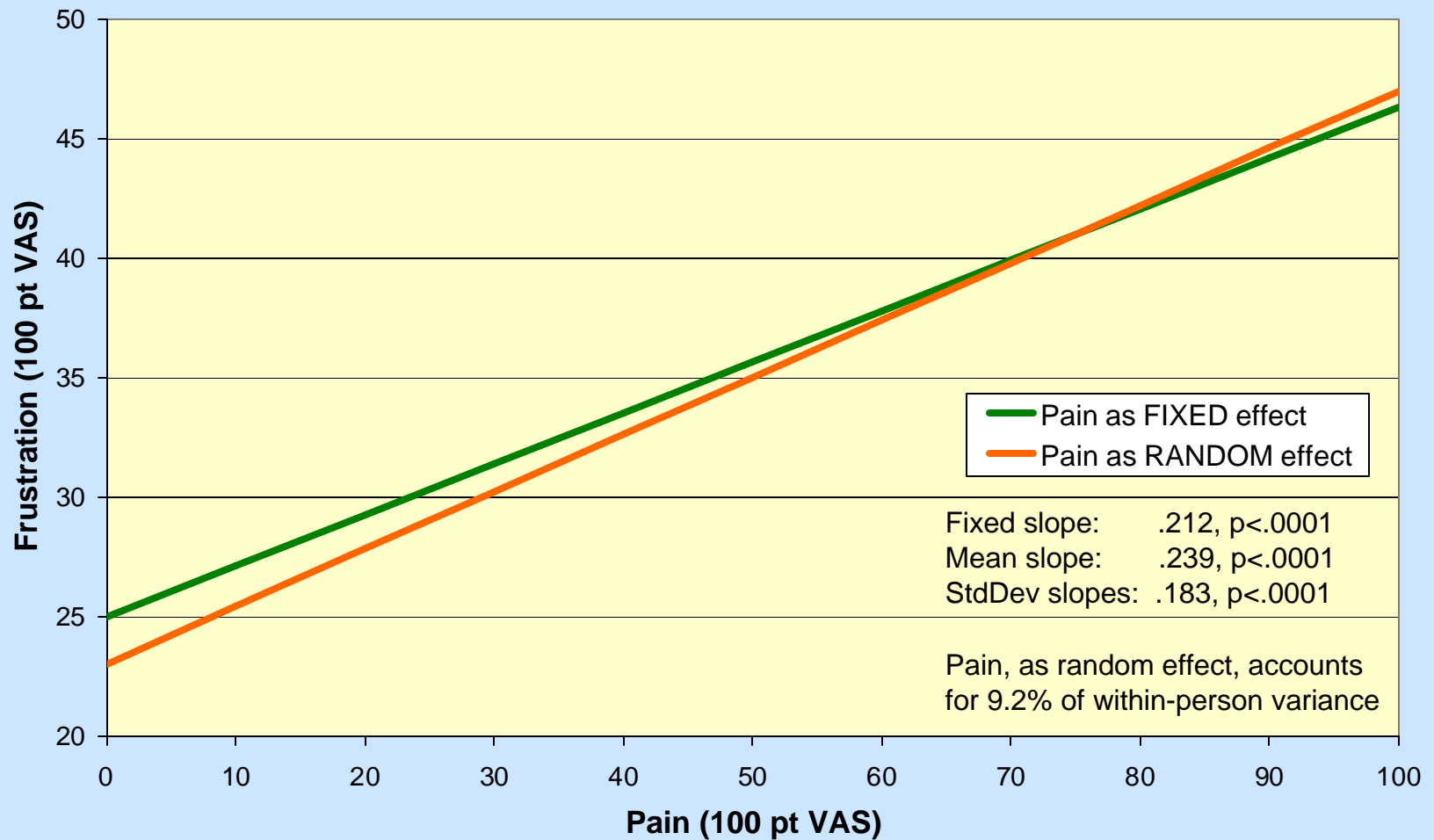


Q4b: Does Relationship of Frustration to Pain Vary Across Persons?

Treating a within-person relationship as a
“random effect”

Provides estimates of
the variance of the person-specific slopes and
the mean (across persons) of these slopes

Within-Person Relationship of Frustration to Pain



Answer to Q4

Yes, changes in pain are positively associated with changes in frustration.

Furthermore, there are substantial individual differences in the slope of this relationship: assuming a normal distribution 67% of slopes are between .056 and .422.

Issue of Causality

Demonstrating that change in one variable is associated with change in a second variable is much stronger evidence for a causal relationship than is a cross-sectional association.

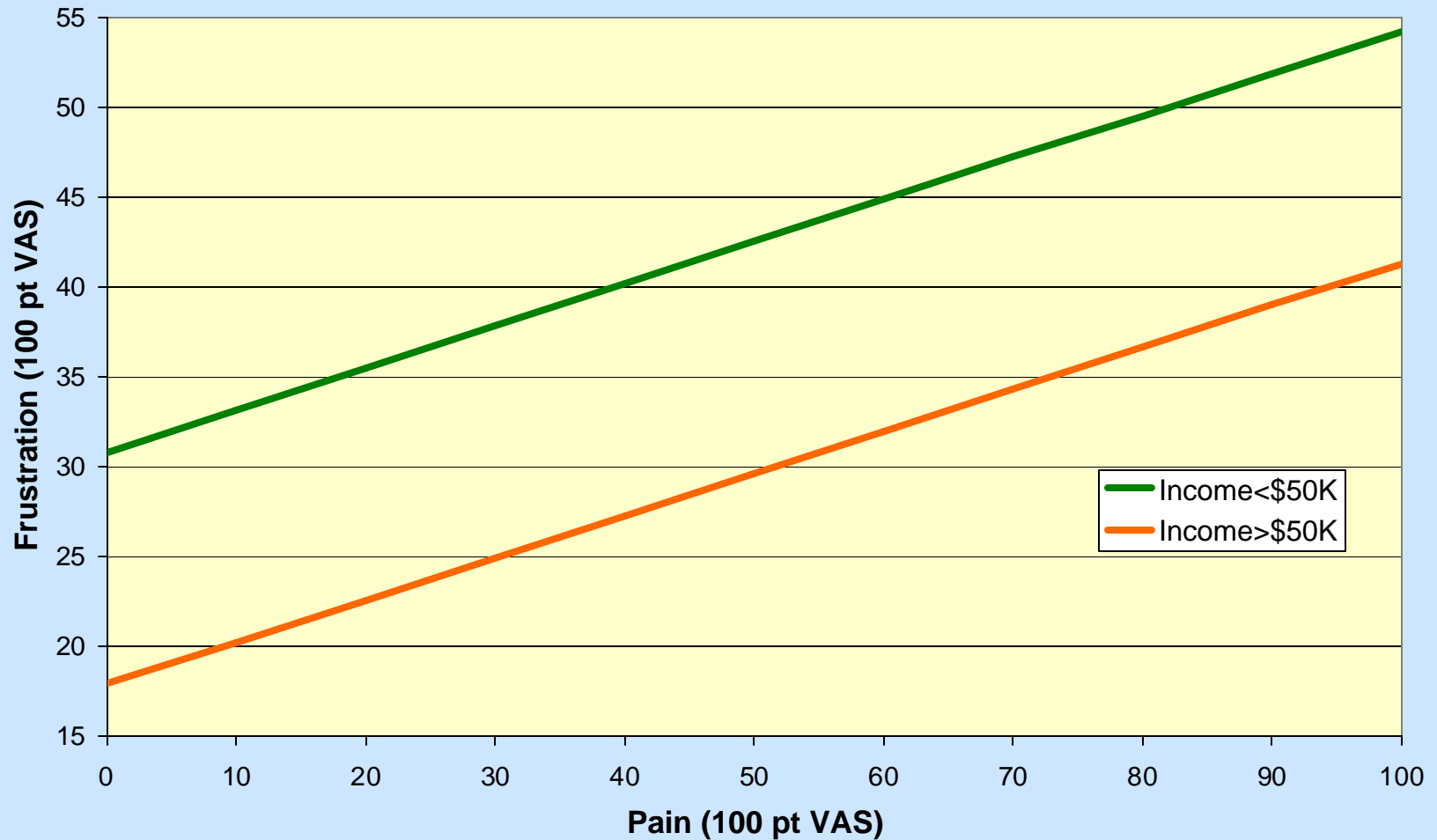
However, it does not establish the direction of the causal relationship, or rule out the possibility that some third variable is causing the change in both variables.

Q5: Are Differences in the Relationship of Frustration to Pain Related to SES? (a between-person factor predicting individual differences in a within-person relationship)

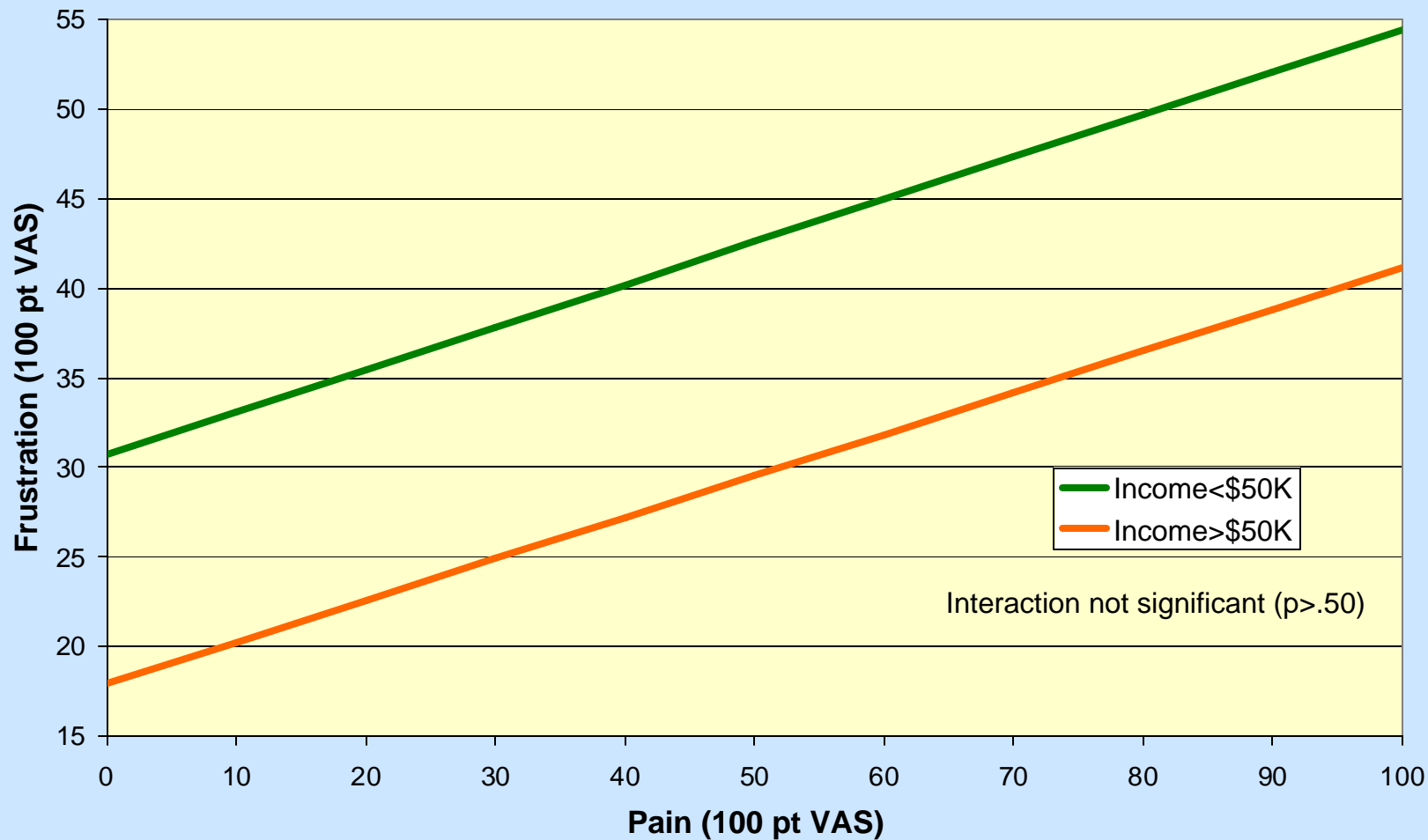
Statistically, this is an interaction effect of a between-person factor (SES) with a within-person factor (pain) in the equation predicting frustration.

Conceptually, we are interested in the extent to which SES may account for the already observed variability among persons in the pain-frustration relationship

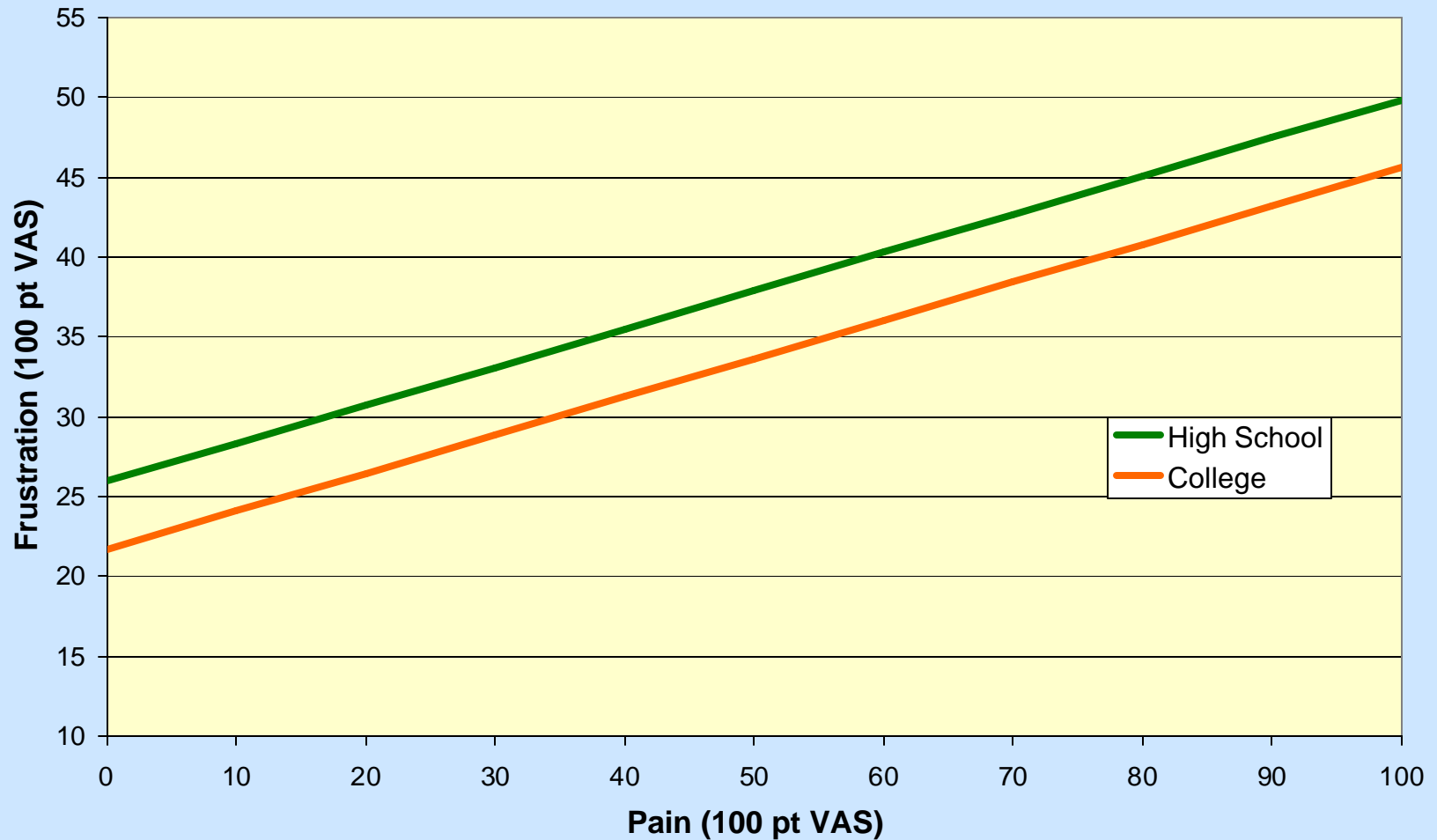
Within-Person Relationship of Frustration to Pain and Family Income Group (additive model)



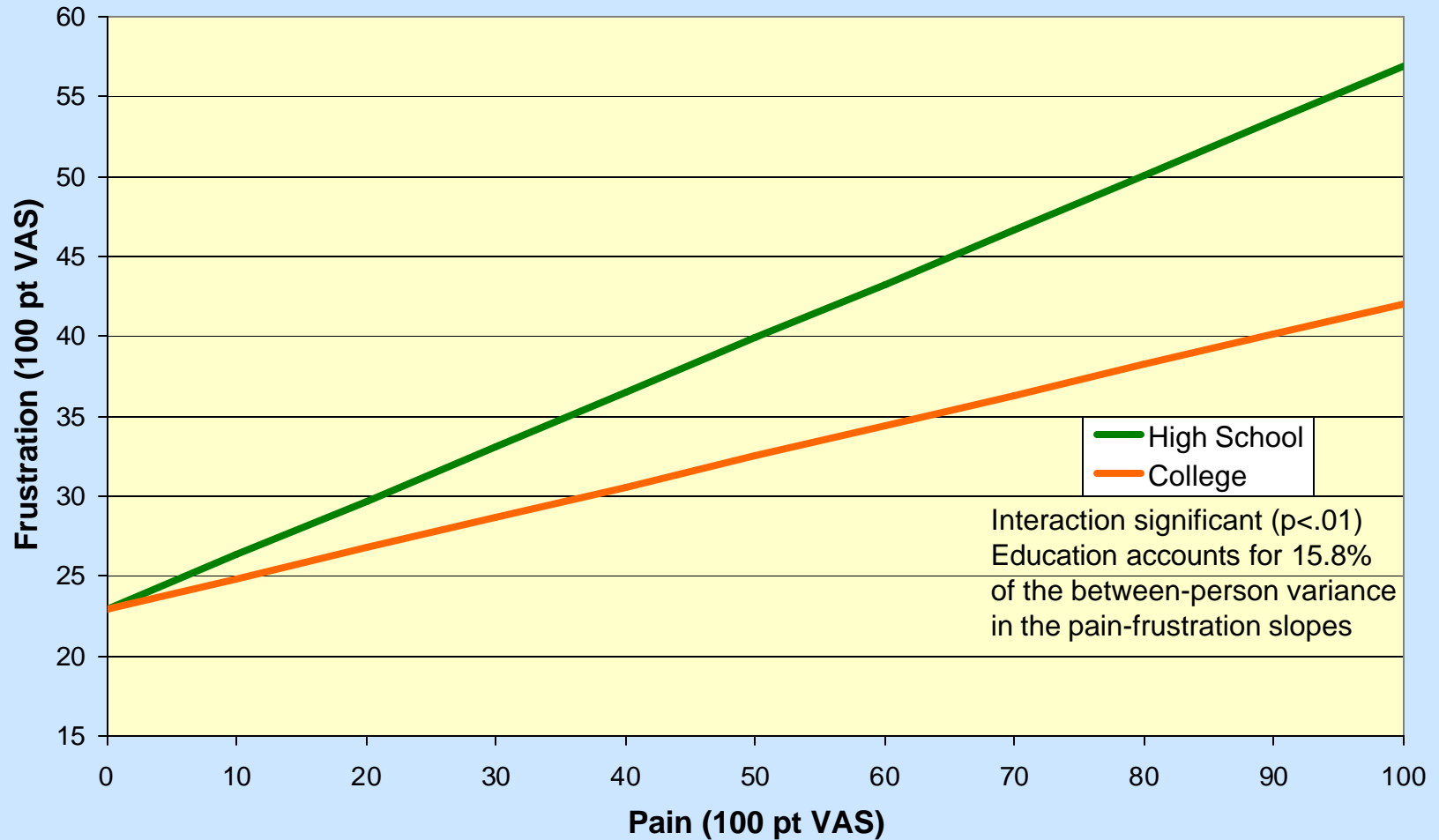
Within-Person Relationship of Frustration to Pain, by Family Income Group (interaction model)



Within-Person Relationship of Frustration to Pain and Education Group (additive model)



Within-Person Relationship of Frustration to Pain, by Education Group (interaction model)



Answer to Q5

Those with family incomes $> \$50,000$ are, on average, less frustrated than those with lower family incomes ($p < .005$), but income does NOT account for individual differences in the pain-frustration relationship ($p > .90$).

In contrast, the 4-point difference in average level of frustration (controlling for pain) between those with and without some college is NOT significant ($p > .40$), but those with only a high school education become much more frustrated as their pain increases, compared to those who attended college ($p < .01$).

Q6: How Reproducible/Stable Are Individual Differences?

- Week 1/Week 2 test-retest correlation of person means
 - empirical estimates of average weekly frustration
 - OLS estimates
 - EBLUP estimates
 - latent variable (multilevel model) approach

Q6: How Reproducible/Stable Are Individual Differences?

	Observed Means	Latent Variable
Means		
Week 1	34.8	34.8
Week 2	35.5	34.4
StdDev		
Week 1	22.1	21.1
Week 2	22.5	21.5
Test-retest r	.895	.943

Answer to Q6

The cross-sectional differences in average frustration level are VERY stable from one week to the next.

This said, the difference in the test-retest correlations of the observed averages (which contain random measurement error due to sampling of moments) and the latent variable averages indicates that the Week 1 to Week 2 computed CHANGE in average frustration is quite unreliable ($r^2=0.50$).

Q6: How Reproducible/Stable Are Individual Differences?

- Week 1/Week 2 test-retest correlation of person means
 - empirical estimates of average weekly frustration
 - OLS estimates
 - EBLUP estimates
 - latent variable approach
- Could also examine Week 1/Week 2 test-retest correlations of slopes (and intercepts) of frustration/pain relationship

Summary - 1

- A wide range of topics/questions that can be addressed
 - The extent of individual differences in mean level of outcome variable
 - Relative amounts of between-person vs within-person variance (intraclass correlation)
 - Are there person-level factors that account for differences in mean levels?

Summary - 2

- Pattern of serial autocorrelation within persons
 - May provide insight into phenomenon you are studying
 - Making correct assumptions about the residuals increases the validity of the primary analyses
 - Often doesn't, but sometimes dramatically affects results

Summary - 3

- What factors might account for within-person fluctuations in the outcome variable?
 - Are there temporal trends/patterns?
 - Are there other within-person (real-time) factors that predict fluctuations in the outcome?
 - Are these relationships constant across all persons (fixed effect), or does the within-person relationship vary across persons (individual differences, random effect)?
 - If the latter, are these individual differences related to measured person-level factors?
- How reproducible are individual differences in
 - mean levels of the outcome
 - within-person relationships with the outcome

Summary - 4

- Issues to think about in multilevel modeling
 - Deciding whether to treat within-person predictors as fixed or random
 - I usually treat within-person covariates (control variables) as fixed factors
 - more parsimonious
 - facilitates estimation of model
 - facilitates interpretation of the model
 - When estimating the interaction effect of a person-level factor with a within-person factor, it is logical to treat the latter's main effect as random

Challenges

- Causality of within-person relationships
 - lagged relationships (appropriate lag period?)
- Different methods of estimation
 - maximum likelihood
 - restricted maximum likelihood
 - general estimating equations
- Creating empirical estimates of latent variables
- With random effects (other than intercept), need to be VERY careful in interpreting intercept and its variance (no longer “between-subject” variance)

Finally

I hope this tale of frustration
has not been too frustrating

Collaborators on analyses involving multilevel modeling of Real-Time data

Joan Broderick

Linda Gerber

Kazuomi Kario

Tessa Pollard

Arthur Stone

Andrew Waters

Elizabeth Brondolo

Tom Kamarck

Kathleen O'Connell

Saul Shiffman

Tom Pickering